William Stallings
Data and Computer Communications
7th Edition

Chapter 18
Internet Protocols

Protocol Functions

- Small set of functions that form basis of all protocols
- Not all protocols have all functions
  - Reduce duplication of effort
  - May have same type of function in protocols at different levels
- Encapsulation
- Fragmentation and reassembly
- Connection control
- Ordered delivery
- Flow control
- Error control
- Addressing
- Multiplexing
- Transmission services

Encapsulation

- Data usually transferred in blocks
  - Protocol data units (PDUs)
  - Each PDU contains data and control information
  - Some PDUs only control
- Three categories of control
- Address
  - Of sender and/or receiver
- Error-detecting code
  - E.g. frame check sequence
- Protocol control
  - Additional information to implement protocol functions
- Addition of control information to data is encapsulation
- Data accepted or generated by entity and encapsulated into PDU
  - Containing data plus control information
  - e.g. TFTP, HDLC, frame relay, ATM, AAL5 (Figure 11.15), LLC, IEEE 802.3, IEEE 802.11

Fragmentation and Reassembly (Segmentation – OSI)

- Exchange data between two entities
- Characterized as sequence of PDUs of some bounded size
  - Application level message
- Lower-level protocols may need to break data up into smaller blocks
- Communications network may only accept blocks of up to a certain size
  - ATM 53 octets
  - Ethernet 1526 octets
- More efficient error control
  - Smaller retransmission
- Fairer
  - Prevent station monopolizing medium
- Smaller buffers
- Provision of checkpoint and restart/recovery operations

Disadvantages of Fragmentation

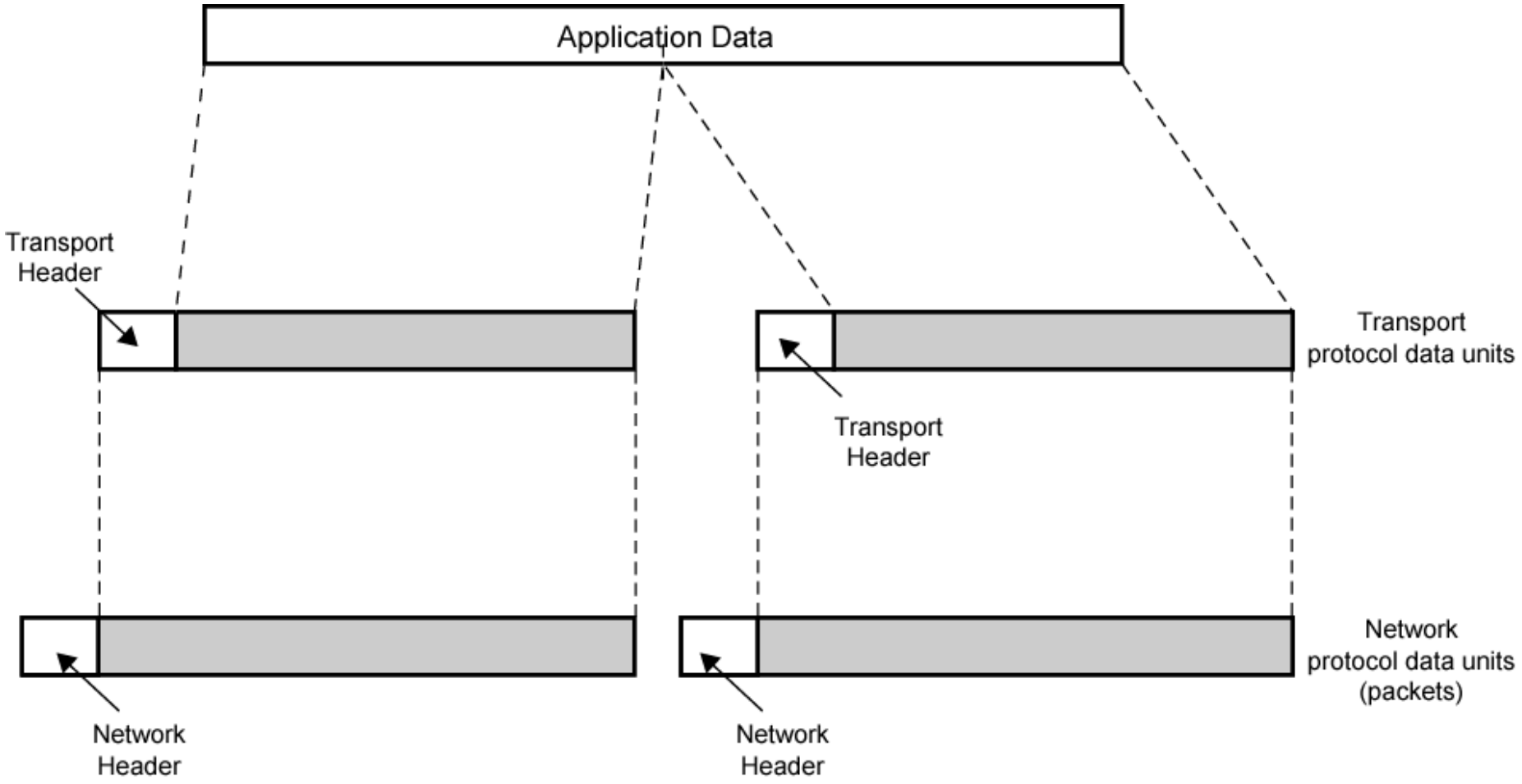
- Make PDUs as large as possible because
  - PDU contains some control information
  - Smaller block, larger overhead
- PDU arrival generates interrupt
  - Smaller blocks, more interrupts
- More time processing smaller, more numerous PDUs
-

Reassembly

- Segmented data must be reassembled into messages
- More complex if PDUs out of order

PDUS and Fragmentation

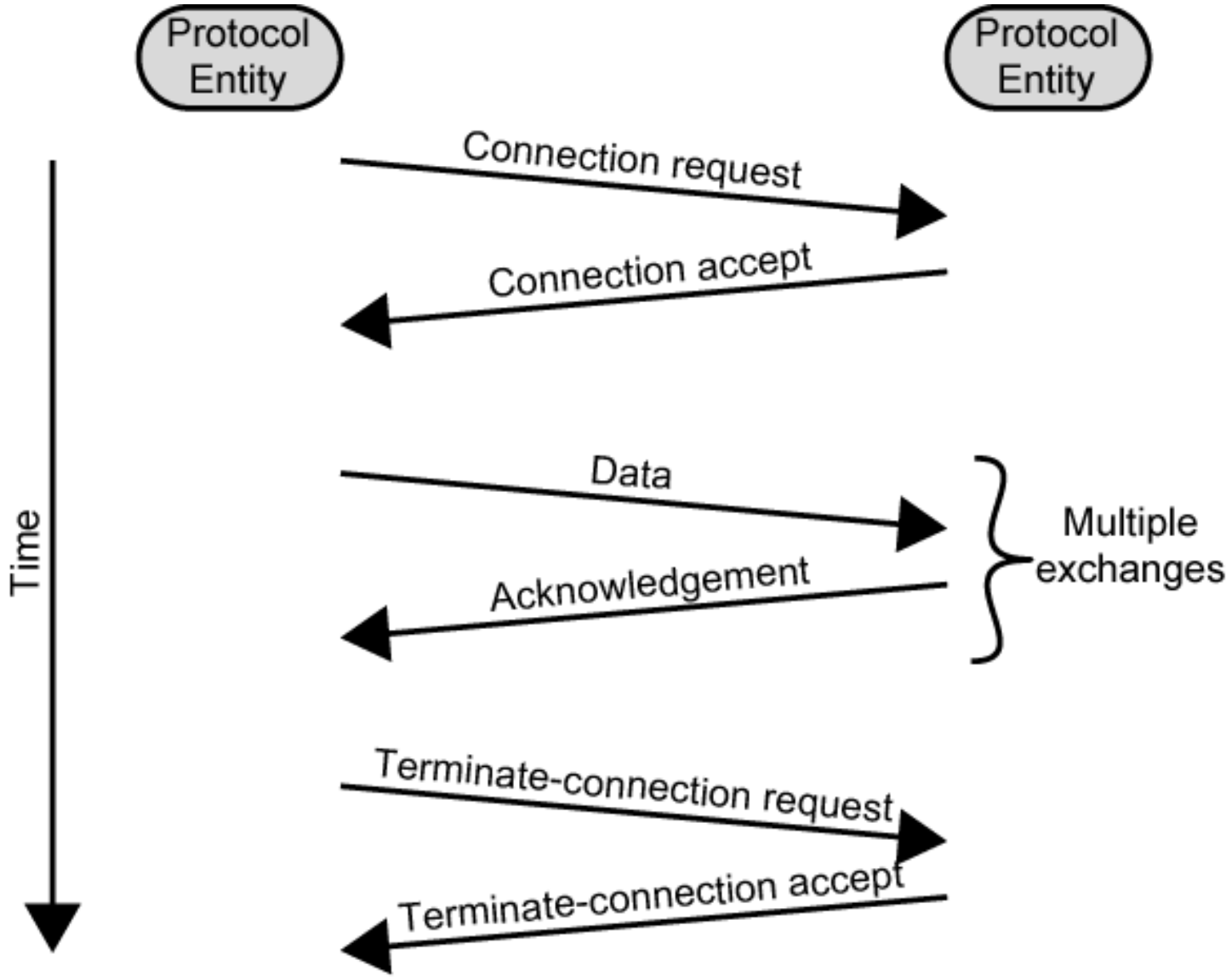
(Copied from chapter 2 fig 2.4)



Connection Control

- Connectionless data transfer
  - Each PDU treated independently
  - E.g. datagram
- Connection-oriented data transfer
  - E.g. virtual circuit
- Connection-oriented preferred (even required) for lengthy exchange of data
- Or if protocol details must be worked out dynamically
- Logical association, or connection, established between entities
- Three phases occur
  - Connection establishment
  - Data transfer
  - Connection termination
  - May be interrupt and recovery phases to handle errors

Phases of Connection Oriented Transfer



Connection Establishment

- Entities agree to exchange data
- Typically, one station issues connection request
  - In connectionless fashion
- May involve central authority
- Receiving entity accepts or rejects (simple)
- May include negotiation
- Syntax, semantics, and timing
- Both entities must use same protocol
- May allow optional features
- Must be agreed
- E.g. protocol may specify max PDU size 8000 octets; one station may wish to restrict to 1000 octets

Data Transfer and Termination

- Both data and control information exchanged
  - —e.g. flow control, error control
- Data flow and acknowledgements may be in one or both directions
- One side may send termination request
- Or central authority might terminate

Sequencing

- Many connection-oriented protocols use sequencing
  - e.g. HDLC, IEEE 802.11
- PDUs numbered sequentially
- Each side keeps track of outgoing and incoming numbers
- Supports three main functions
  - Ordered delivery
  - Flow control
  - Error control
- Not found in all connection-oriented protocols
  - E.g.frame relay and ATM
- All connection-oriented protocols include some way of identifying connection
  - Unique connection identifier
  - Combination of source and destination addresses

Ordered Delivery

- PDUs may arrive out of order
  - Different paths through network
- PDU order must be maintained
- Number PDUs sequentially
- Easy to reorder received PDUs
- Finite sequence number field
  - Numbers repeat modulo maximum number
  - Maximum sequence number greater than maximum number of PDUs that could be outstanding
  - In fact, maximum number may need to be twice maximum number of PDUs that could be outstanding
    - e.g. selective-repeat ARQ

Flow Control

- Performed by receiving entity to limit amount or rate of data sent
- Stop-and-wait
  - Each PDU must be acknowledged before next sent
- Credit
  - Amount of data that can be sent without acknowledgment
  - E.g. HDLC sliding-window
- Must be implemented in several protocols
  - Network traffic control
  - Buffer space
  - Application overflow
    - E.g. waiting for disk access

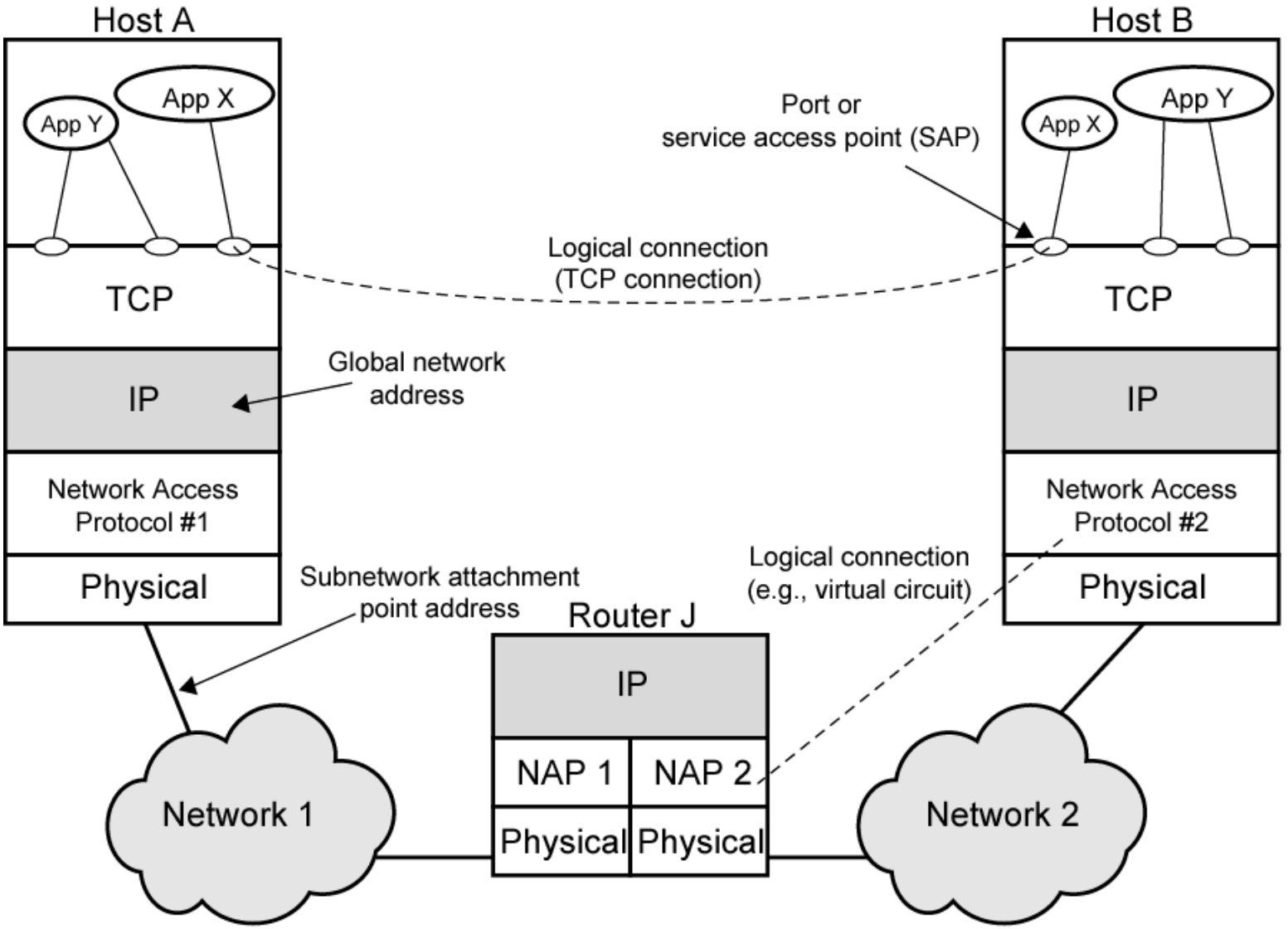
Error Control

- Guard against loss or damage
- Error detection and retransmission
  - Sender inserts error-detecting code in PDU
    - Function of other bits in PDU
  - Receiver checks code on incoming PDU
  - If error, discard
  - If transmitter doesn't get acknowledgment in reasonable time, retransmit
- Error-correction code
- Enables receiver to detect and possibly correct errors
- Error control is performed at various layers of protocol
  - Between station and network
  - Inside network

Addressing

- Addressing level
- Addressing scope
- Connection identifiers
- Addressing mode

TCP/IP Concepts



Addressing Level

- Level in comms architecture at which entity is named
- Unique address for each end system
  - e.g. workstation or server
- And each intermediate system
  - (e.g., router)
- Network-level address
  - IP address or internet address
  - OSI - network service access point (NSAP)
  - Used to route PDU through network
- At destination data must routed to some process
  - Each process assigned an identifier
  - TCP/IP port
  - Service access point (SAP) in OSI

Addressing Scope

- Global address
  - Global nonambiguity
  - Identifies unique system
  - Synonyms permitted
  - System may have more than one global address
  - Global applicability
  - Possible at any global address to identify any other global address, in any system, by means of global address of other system
  - Enables internet to route data between any two systems
- Need unique address for each device interface on network
  - MAC address on IEEE 802 network and ATM host address
  - Enables network to route data units through network and deliver to intended system
  - Network attachment point address
- Addressing scope only relevant for network-level addresses
- Port or SAP above network level is unique within system
  - Need not be globally unique
  - E.g port 80 web server listening port in TCP/IP

Connection Identifiers

- Entity 1 on system A requests connection to entity 2 on system B, using global address B.2.
- B.2 accepts connection
- Connection identifier used by both entities for future transmissions
- Reduced overhead
  - Generally shorter than global identifiers
- Routing
  - Fixed route may be defined
  - Connection identifier identifies route to intermediate systems
- Multiplexing
  - Entity may wish more than one connection simultaneously
  - PDUs must be identified by connection identifier
- Use of state information
- Once connection established, end systems can maintain state information about connection
  - Flow and error control using sequence numbers

Addressing Mode

- Usually address refers to single system or port
  - —Individual or unicast address
- Address can refer to more than one entity or port
  - —Multiple simultaneous recipients for data
  - —Broadcast for all entities within domain
  - —Multicast for specific subset of entities

Multiplexing

- Multiple connections into single system
  - —E.g. frame relay, can have multiple data link connections terminating in single end system
  - —Connections multiplexed over single physical interface
- Can also be accomplished via port names
  - —Also permit multiple simultaneous connections
  - —E.g. multiple TCP connections to given system
    - Each connection on different pair of ports

Multiplexing Between Levels

- Upward or inward multiplexing
  - —Multiple higher-level connections share single lower-level connection
    - More efficient use of lower-level service
    - Provides several higher-level connections where only single lower-level connection exists
- Downward multiplexing, or splitting
  - —Higher-level connection built on top of multiple lower-level connections
  - —Traffic on higher connection divided among lower connections
    - Reliability, performance, or efficiency.

Transmission Services

- Protocol may provide additional services to entities
- E.g.:
- Priority
  - Connection basis
  - On message basis
    - E.g. terminate-connection request
- Quality of service
  - E.g. minimum throughput or maximum delay threshold
- Security
  - Security mechanisms, restricting access
- These services depend on underlying transmission system and lower-level entities

Internetworking Terms (1)

- Communications Network
  - Facility that provides data transfer service
- An internet
  - Collection of communications networks interconnected by bridges and/or routers
- The Internet - note upper case I
  - The global collection of thousands of individual machines and networks
- Intranet
  - Corporate internet operating within the organization
  - Uses Internet (TCP/IP and http)technology to deliver documents and resources

Internetworking Terms (2)

- End System (ES)
  - —Device attached to one of the networks of an internet
  - —Supports end-user applications or services
- Intermediate System (IS)
  - —Device used to connect two networks
  - —Permits communication between end systems attached to different networks

Internetworking Terms (3)

- Bridge
  - —IS used to connect two LANs using similar LAN protocols
  - —Address filter passing on packets to the required network only
  - —OSI layer 2 (Data Link)
- Router
  - —Connects two (possibly dissimilar) networks
  - —Uses internet protocol present in each router and end system
  - —OSI Layer 3 (Network)

Requirements of Internetworking

- Link between networks
  - —Minimum physical and link layer
- Routing and delivery of data between processes on different networks
- Accounting services and status info
- Independent of network architectures

Network Architecture Features

- Addressing
- Packet size
- Access mechanism
- Timeouts
- Error recovery
- Status reporting
- Routing
- User access control
- Connection based or connectionless

Architectural Approaches

- Connection oriented
- Connectionless

Connection Oriented

- Assume that each network is connection oriented
- IS connect two or more networks
  - —IS appear as ES to each network
  - —Logical connection set up between ESs
    - Concatenation of logical connections across networks
  - —Individual network virtual circuits joined by IS
- May require enhancement of local network services
  - —802, FDDI are datagram services

Connection Oriented IS Functions

- Relaying
- Routing

- e.g. X.75 used to interconnect X.25 packet switched networks

- Connection oriented not often used
  - (IP dominant)

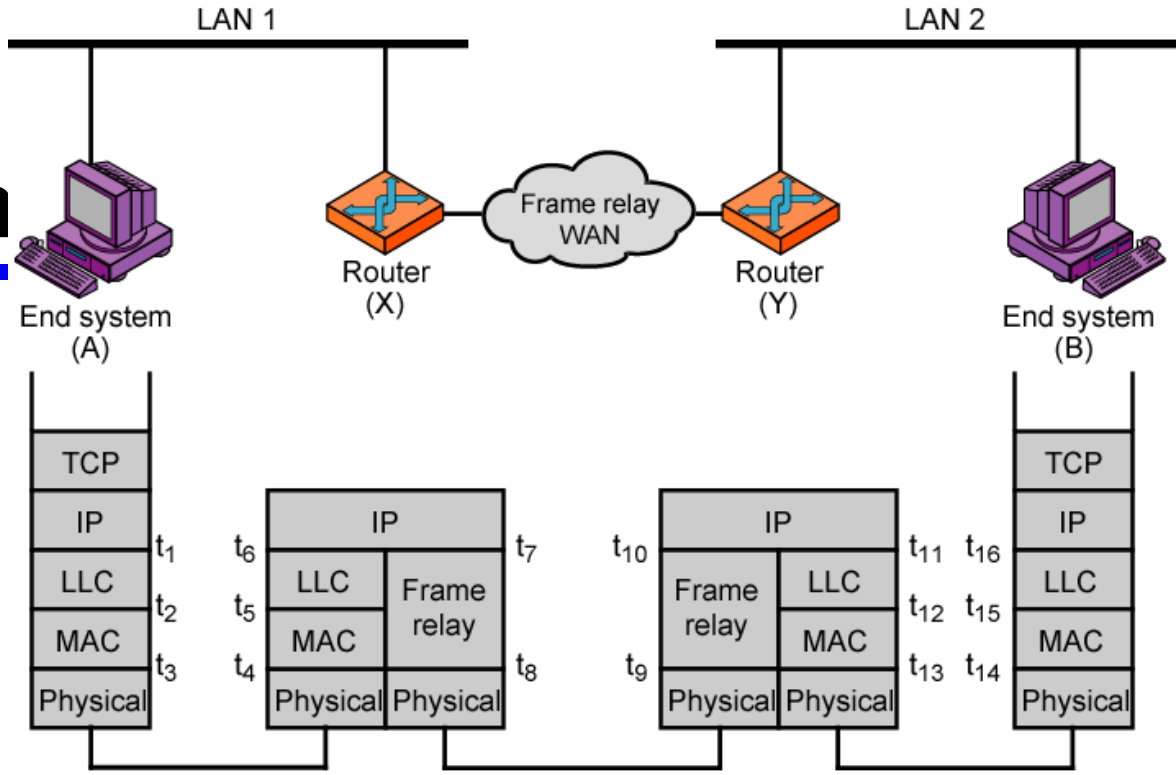
Connectionless Operation

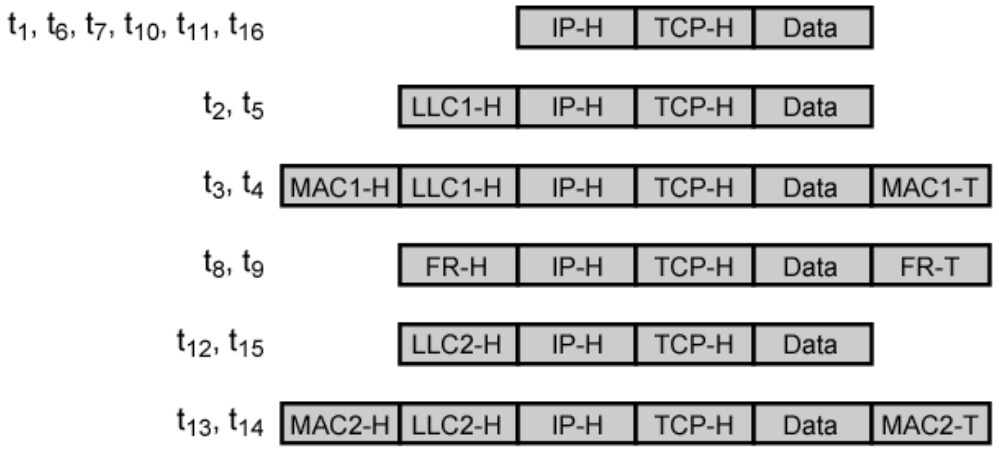
- Corresponds to datagram mechanism in packet switched network
- Each NPDU treated separately
- Network layer protocol common to all DTEs and routers
  - Known generically as the internet protocol
- Internet Protocol
  - One such internet protocol developed for ARPANET
  - RFC 791 (Get it and study it)
- Lower layer protocol needed to access particular network

Connectionless Internetworking

- Advantages
  - —Flexibility
  - —Robust
  - —No unnecessary overhead
- Unreliable
  - —Not guaranteed delivery
  - —Not guaranteed order of delivery
    - Packets can take different routes
  - —Reliability is responsibility of next layer up (e.g. TCP)

IP Operation



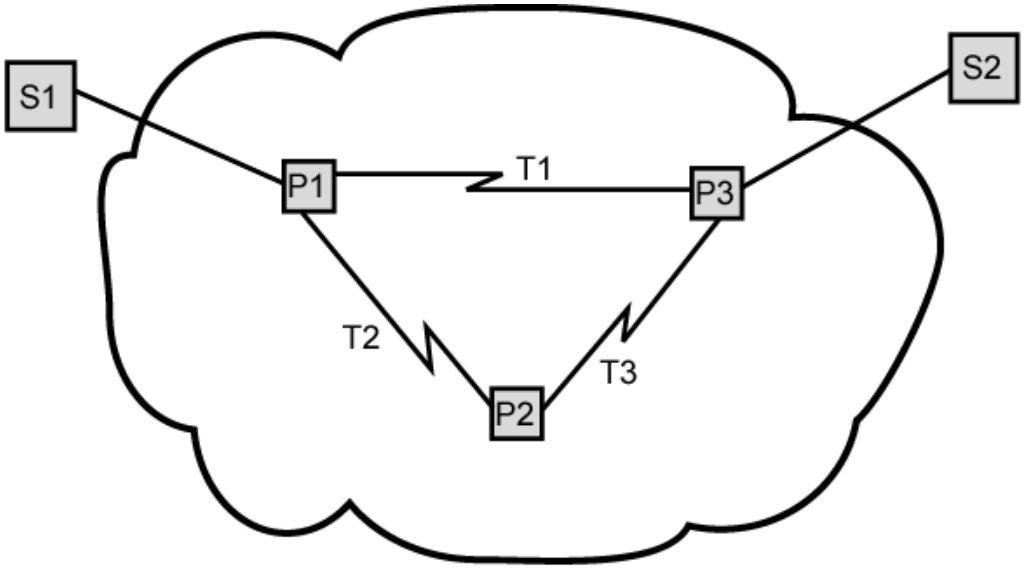


TCP-H = TCP header
IP-H = IP header
LLCi-H = LLC header
MACi-H = MAC header
MACi-T = MAC trailer
FR-H = Frame relay header
FR-T = Frame relay trailer

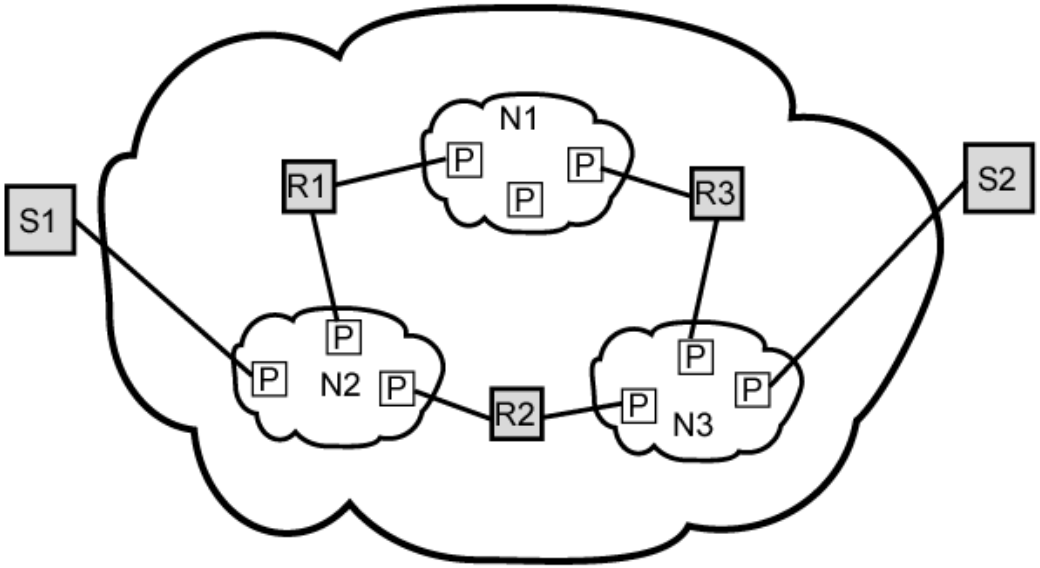
Design Issues

- Routing
- Datagram lifetime
- Fragmentation and re-assembly
- Error control
- Flow control

The Internet as a Network



(a) Packet-switching network architecture



(b) Internetwork architecture

Routing

- End systems and routers maintain routing tables
  - Indicate next router to which datagram should be sent
  - Static
    - May contain alternative routes
  - Dynamic
    - Flexible response to congestion and errors
- Source routing
  - Source specifies route as sequential list of routers to be followed
  - Security
  - Priority
- Route recording

Datagram Lifetime

- Datagrams could loop indefinitely
  - Consumes resources
  - Transport protocol may need upper bound on datagram life
- Datagram marked with lifetime
  - Time To Live field in IP
  - Once lifetime expires, datagram discarded (not forwarded)
  - Hop count
    - Decrement time to live on passing through a each router
  - Time count
    - Need to know how long since last router
- (Aside: compare with Logan's Run)

Fragmentation and Re-assembly

- Different packet sizes
- When to re-assemble
  - —At destination
    - Results in packets getting smaller as data traverses internet
  - —Intermediate re-assembly
    - Need large buffers at routers
    - Buffers may fill with fragments
    - All fragments must go through same router
      - Inhibits dynamic routing

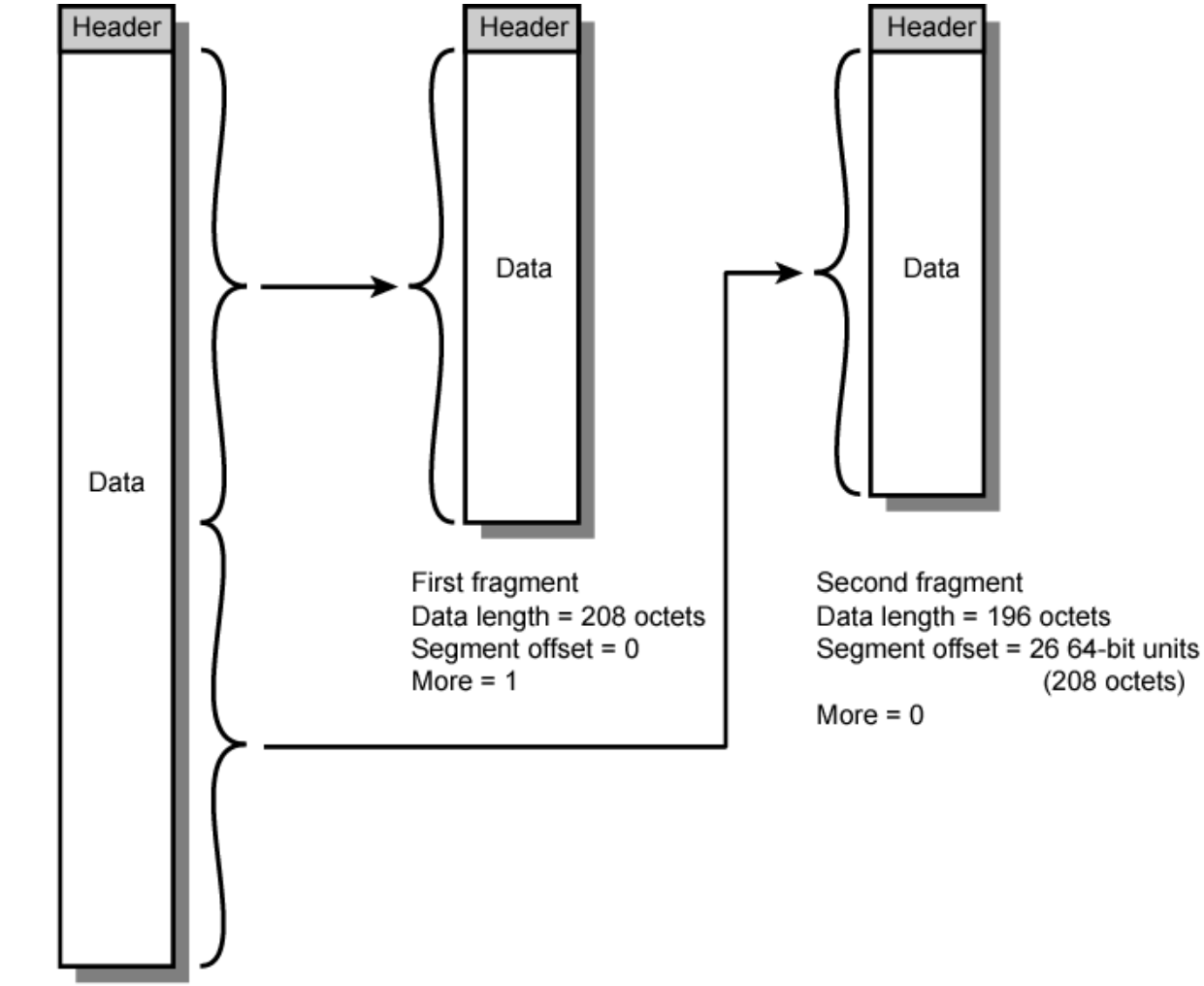
IP Fragmentation (1)

- IP re-assembles at destination only
- Uses fields in header
  - —Data Unit Identifier (ID)
    - Identifies end system originated datagram
      - Source and destination address
      - Protocol layer generating data (e.g. TCP)
      - Identification supplied by that layer
  - —Data length
    - Length of user data in octets

IP Fragmentation (2)

- Offset
  - Position of fragment of user data in original datagram
  - In multiples of 64 bits (8 octets)
- *More* flag
  - Indicates that this is not the last fragment

Fragmentation Example



Dealing with Failure

- Re-assembly may fail if some fragments get lost
- Need to detect failure
- Re-assembly time out
  - —Assigned to first fragment to arrive
  - —If timeout expires before all fragments arrive, discard partial data
- Use packet lifetime (time to live in IP)
  - —If time to live runs out, kill partial data

Error Control

- Not guaranteed delivery
- Router should attempt to inform source if packet discarded
  - e.g. for time to live expiring
- Source may modify transmission strategy
- May inform high layer protocol
- Datagram identification needed
- (Look up ICMP)

Flow Control

- Allows routers and/or stations to limit rate of incoming data
- Limited in connectionless systems
- Send flow control packets
  - Requesting reduced flow
- e.g. ICMP

Internet Protocol (IP) Version 4

- Part of TCP/IP
  - —Used by the Internet
- Specifies interface with higher layer
  - —e.g. TCP
- Specifies protocol format and mechanisms
- RFC 791
  - —Get it and study it!
  - —www.rfc-editor.org
- Will (eventually) be replaced by IPv6 (see later)

IP Services

- Primitives
  - —Functions to be performed
  - —Form of primitive implementation dependent
    - e.g. subroutine call
  - —Send
    - Request transmission of data unit
  - —Deliver
    - Notify user of arrival of data unit
- Parameters
  - —Used to pass data and control info

Parameters (1)

- Source address
- Destination address
- Protocol
  - Recipient e.g. TCP
- Type of Service
  - Specify treatment of data unit during transmission through networks
- Identification
  - Source, destination address and user protocol
  - Uniquely identifies PDU
  - Needed for re-assembly and error reporting
  - Send only

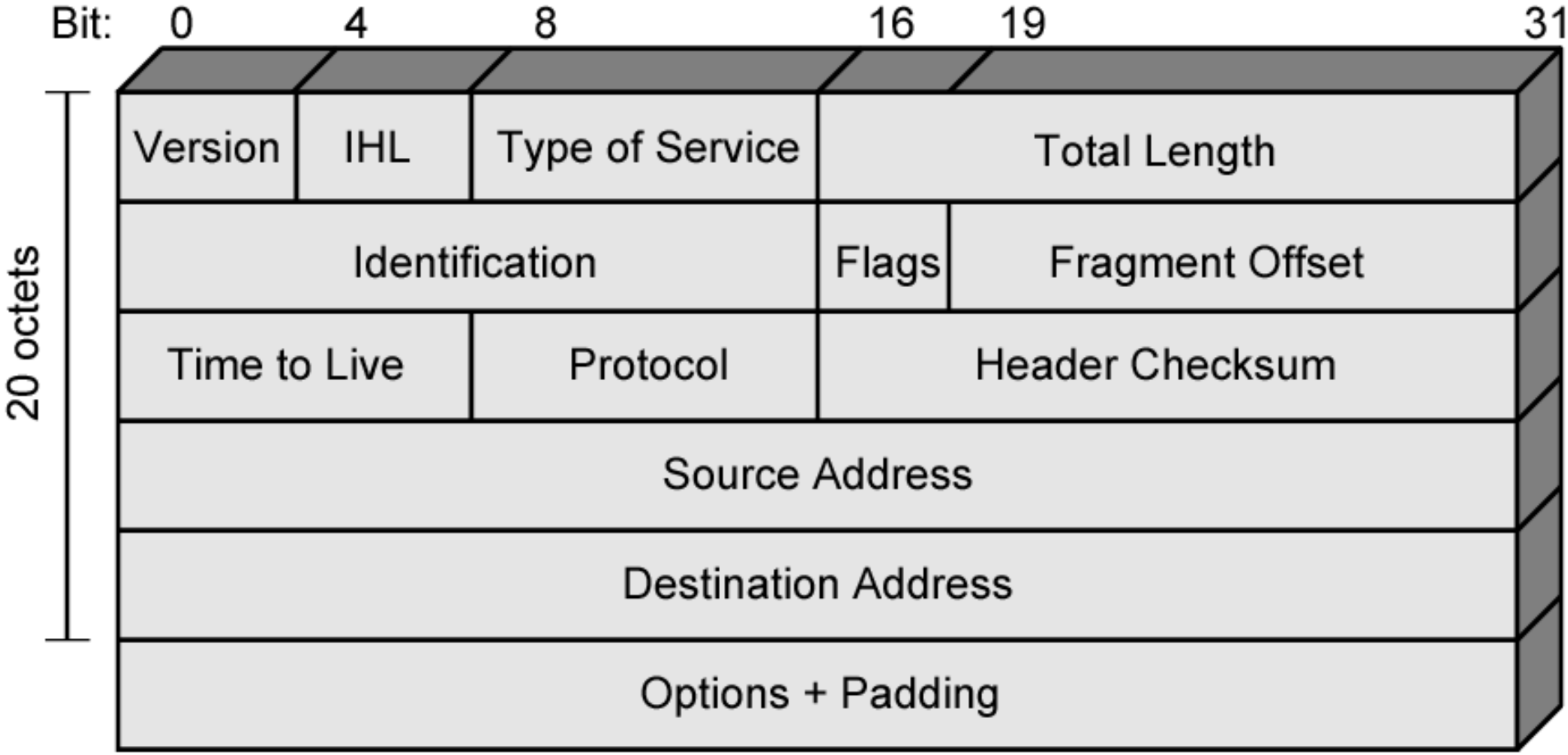
Parameters (2)

- Don’t fragment indicator
  - —Can IP fragment data
  - —If not, may not be possible to deliver
  - —Send only
- Time to live
  - —Send only
- Data length
- Option data
- User data

Options

- Security
- Source routing
- Route recording
- Stream identification
- Timestamping

IPv4 Header


Bit:
0
4
8
16
19
31
Version
IHL
Type of Service
Total Length
Identification
Flags
Fragment Offset
Time to Live
Protocol
Header Checksum
Source Address
Destination Address
Options + Padding
20 octets

Header Fields (1)

- Version
  - —Currently 4
  - —IP v6 - see later
- Internet header length
  - —In 32 bit words
  - —Including options
- Type of service
- Total length
  - —Of datagram, in octets

Header Fields (2)

- Identification
  - Sequence number
  - Used with addresses and user protocol to identify datagram uniquely
- Flags
  - More bit
  - Don’t fragment
- Fragmentation offset
- Time to live
- Protocol
  - Next higher layer to receive data field at destination

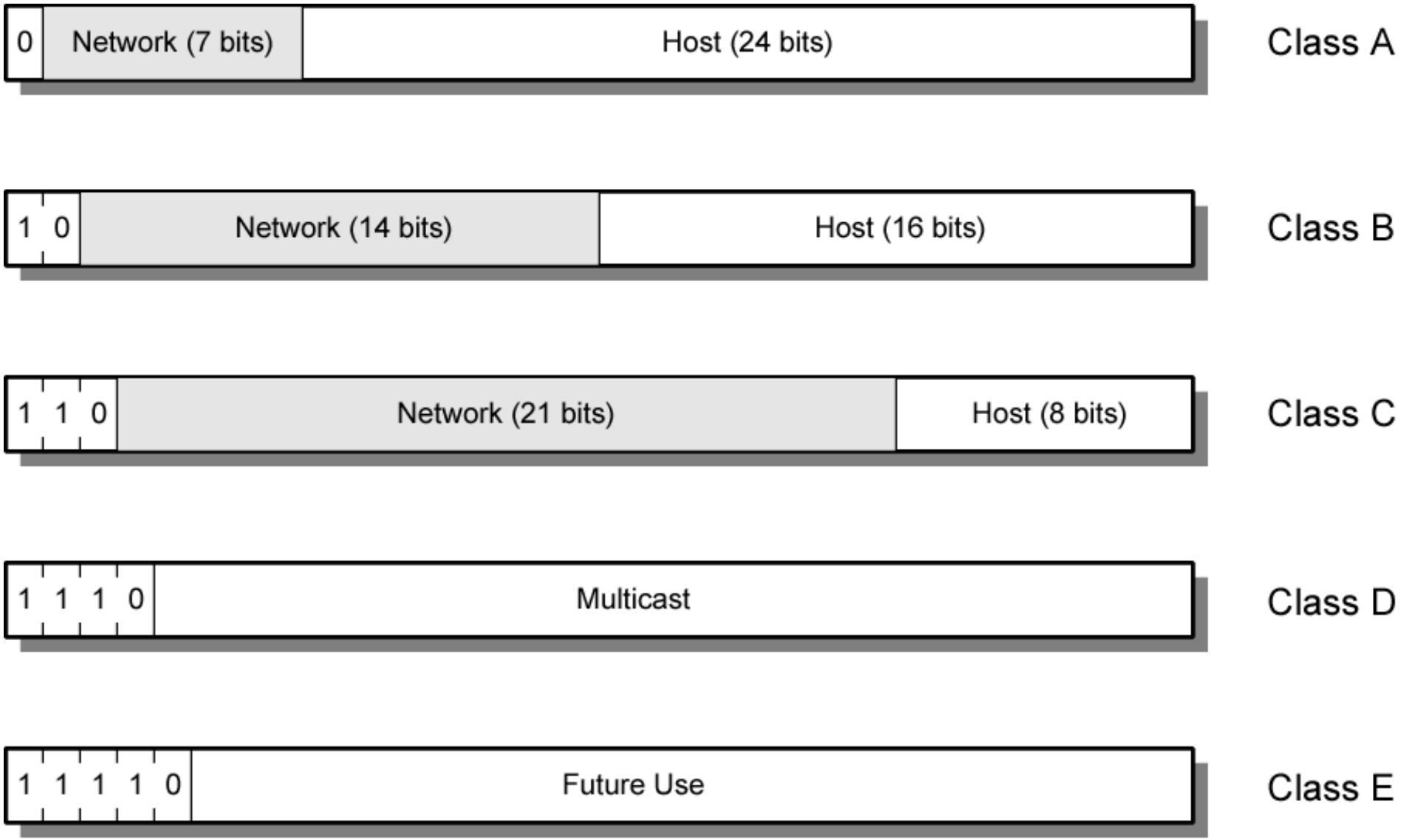
Header Fields (3)

- Header checksum
  - —Reverified and recomputed at each router
  - —16 bit ones complement sum of all 16 bit words in header
  - —Set to zero during calculation
- Source address
- Destination address
- Options
- Padding
  - —To fill to multiple of 32 bits long

Data Field

- Carries user data from next layer up
- Integer multiple of 8 bits long (octet)
- Max length of datagram (header plus data) 65,535 octets

IPv4 Address Formats



IP Addresses - Class A

- 32 bit global internet address
- Network part and host part
- Class A
  - —Start with binary 0
  - —All 0 reserved
  - —01111111 (127) reserved for loopback
  - —Range 1.x.x.x to 126.x.x.x
  - —All allocated

IP Addresses - Class B

- Start 10
- Range 128.x.x.x to 191.x.x.x
- Second Octet also included in network address
- 2^{14} = 16,384 class B addresses
- All allocated

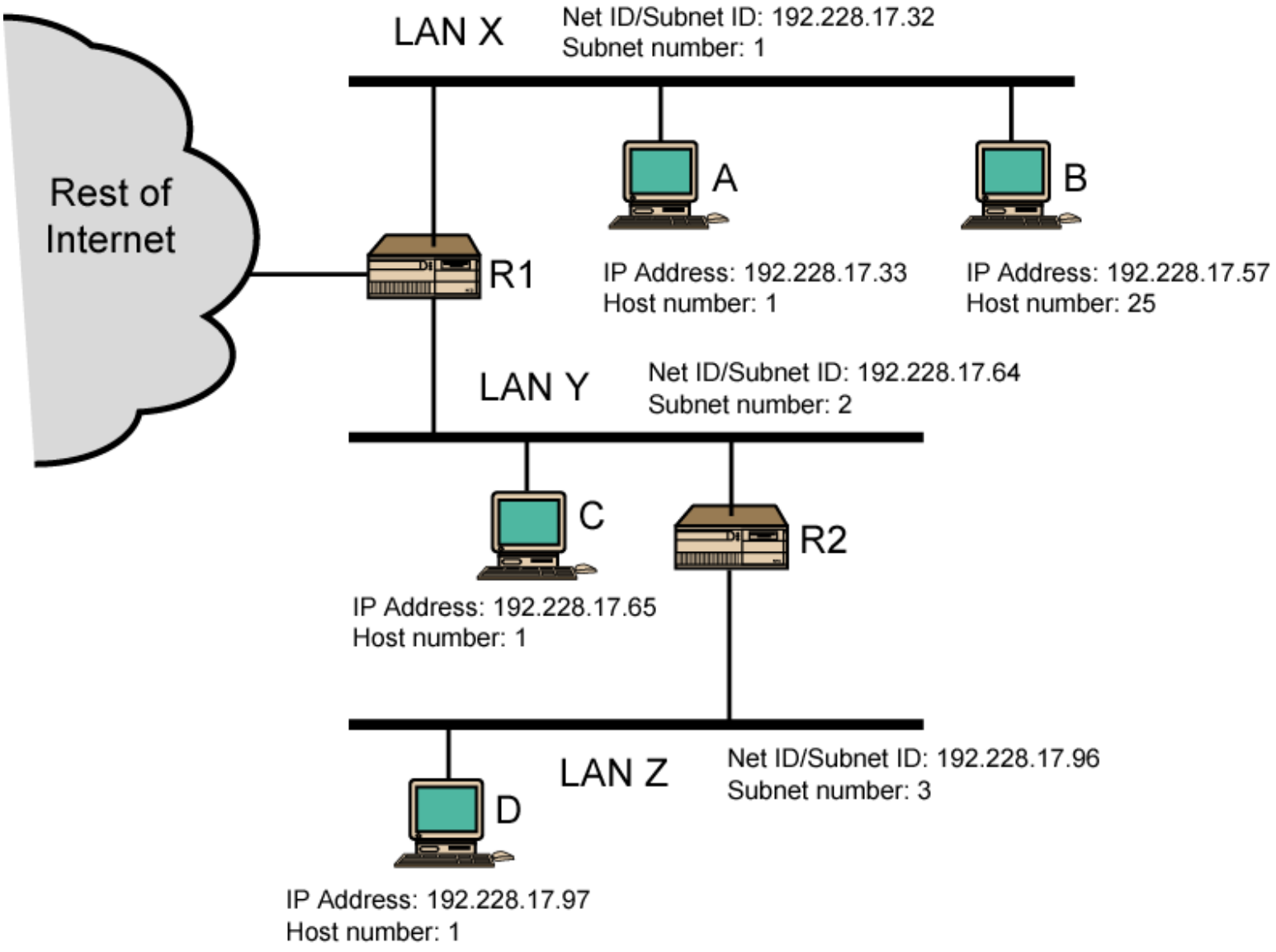
IP Addresses - Class C

- Start 110
- Range 192.x.x.x to 223.x.x.x
- Second and third octet also part of network address
- 2^{21} = 2,097,152 addresses
- Nearly all allocated
  - See IPv6

Subnets and Subnet Masks

- Allow arbitrary complexity of internetworked LANs within organization
- Insulate overall internet from growth of network numbers and routing complexity
- Site looks to rest of internet like single network
- Each LAN assigned subnet number
- Host portion of address partitioned into subnet number and host number
- Local routers route within subnetted network
- Subnet mask indicates which bits are subnet number and which are host number

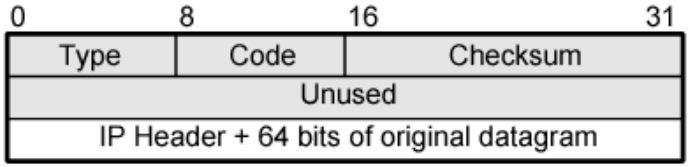
Routing Using Subnets



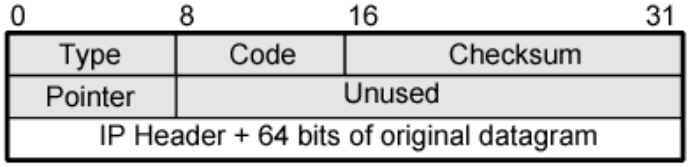
ICMP

- Internet Control Message Protocol
- RFC 792 (get it and study it)
- Transfer of (control) messages from routers and hosts to hosts
- Feedback about problems
  - e.g. time to live expired
- Encapsulated in IP datagram
  - Not reliable

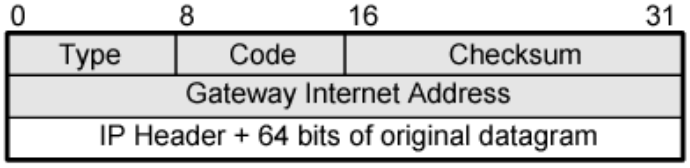
ICMP Message Formats



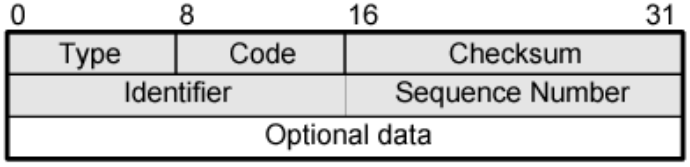
(a) Destination Unreachable; Time Exceeded; Source Quench



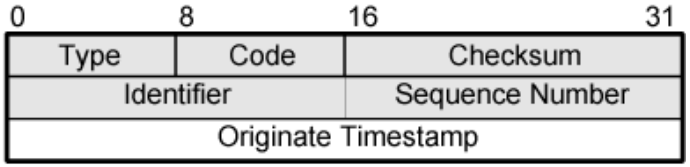
(b) Parameter Problem



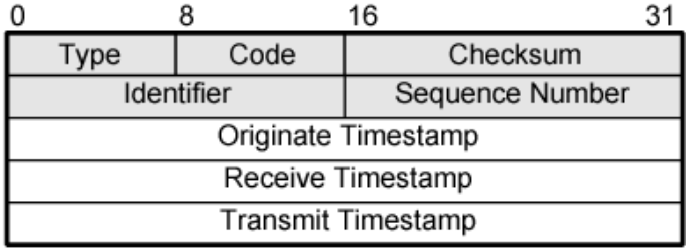
(c) Redirect



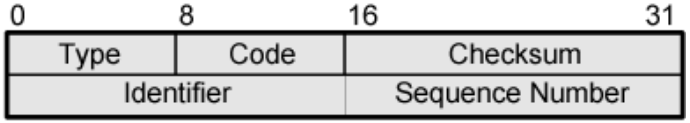
(d) Echo, Echo Reply



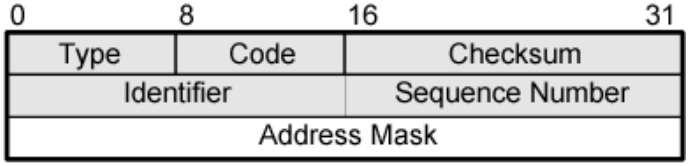
(e) Timestamp



(f) Timestamp Reply



(g) Address Mask Request



(h) Address Mask Reply

IP v6 - Version Number

- IP v 1-3 defined and replaced
- IP v4 - current version
- IP v5 - streams protocol
- IP v6 - replacement for IP v4
  - During development it was called IPng
  - Next Generation

Why Change IP?

- Address space exhaustion
  - —Two level addressing (network and host) wastes space
  - —Network addresses used even if not connected to Internet
  - —Growth of networks and the Internet
  - —Extended use of TCP/IP
  - —Single address per host
- Requirements for new types of service

IPv6 RFCs

- 1752 - Recommendations for the IP Next Generation Protocol
- 2460 - Overall specification
- 2373 - addressing structure
- others (find them)
- www.rfc-editor.org

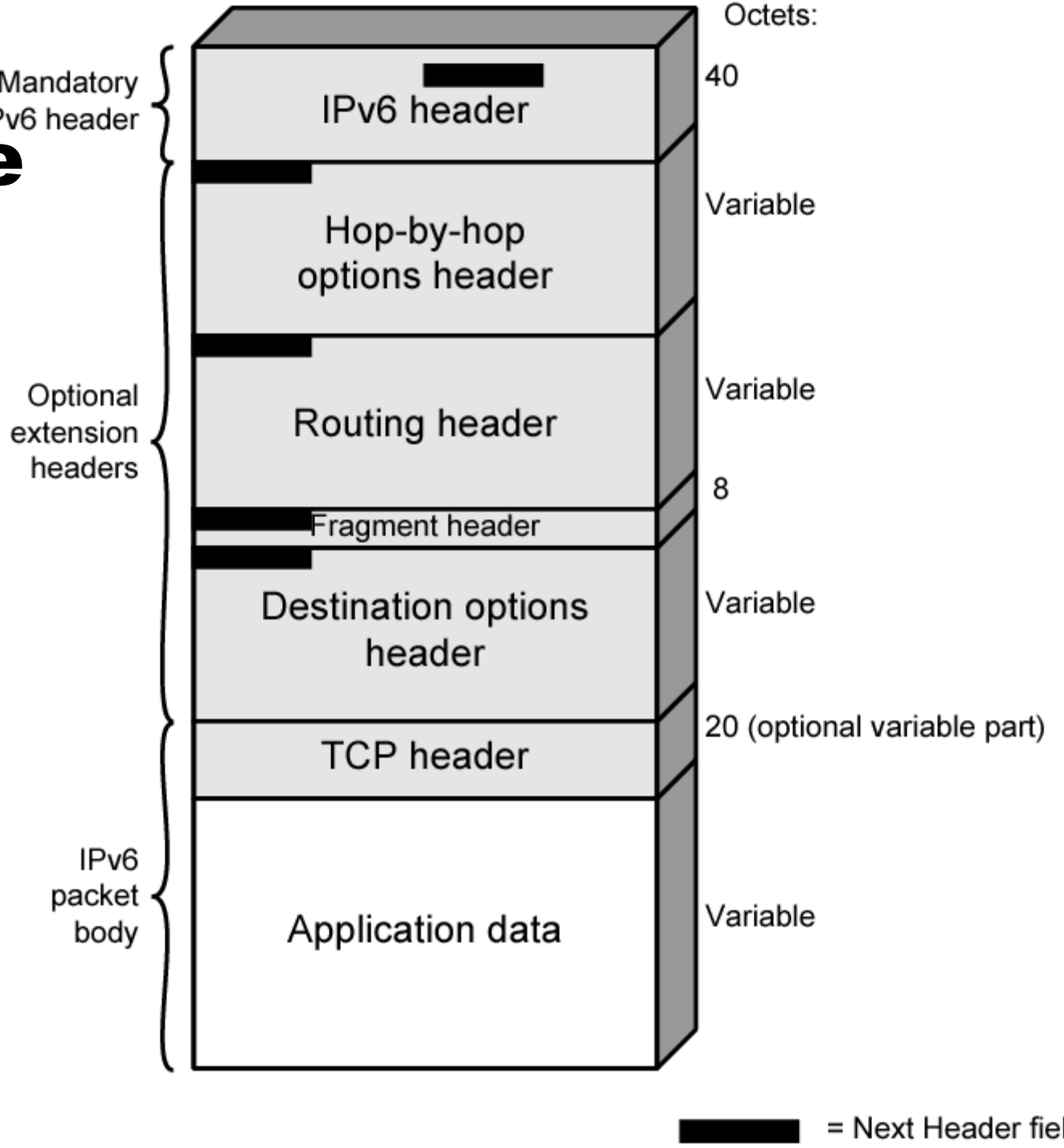
IPv6 Enhancements (1)

- Expanded address space
  - —128 bit
- Improved option mechanism
  - —Separate optional headers between IPv6 header and transport layer header
  - —Most are not examined by intermediate routes
    - Improved speed and simplified router processing
    - Easier to extend options
- Address autoconfiguration
  - —Dynamic assignment of addresses

IPv6 Enhancements (2)

- Increased addressing flexibility
  - —Anycast - delivered to one of a set of nodes
  - —Improved scalability of multicast addresses
- Support for resource allocation
  - —Replaces type of service
  - —Labeling of packets to particular traffic flow
  - —Allows special handling
  - —e.g. real time video

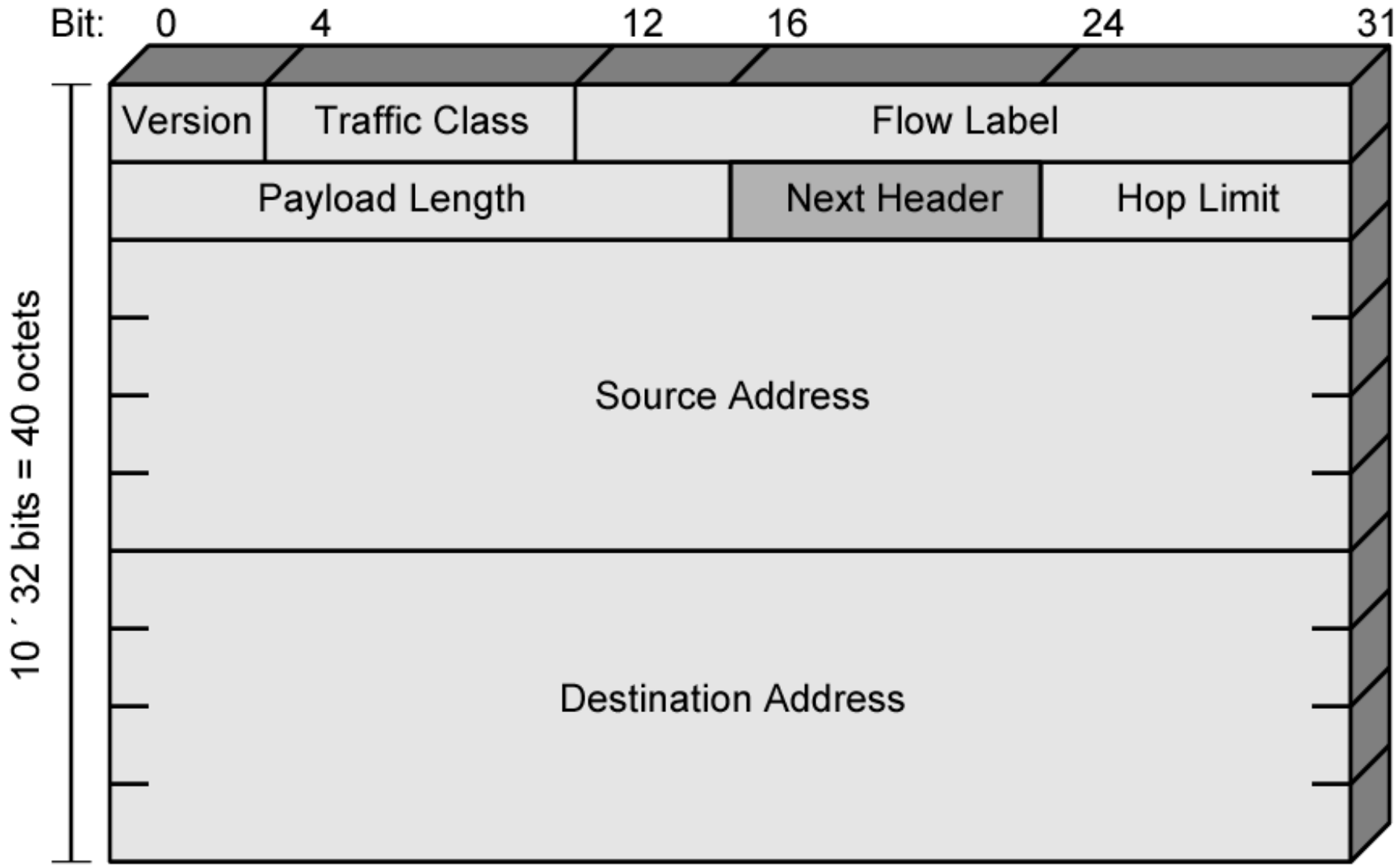
IPv6 Structure



Extension Headers

- Hop-by-Hop Options
  - —Require processing at each router
- Routing
  - —Similar to v4 source routing
- Fragment
- Authentication
- Encapsulating security payload
- Destination options
  - —For destination node

IP v6 Header


Bit:
0
4
12
16
24
31
Version
Traffic Class
Flow Label
Payload Length
Next Header
Hop Limit
Source Address
Destination Address
10 ´ 32 bits = 40 octets

IP v6 Header Fields (1)

- Version
  - —6
- Traffic Class
  - —Classes or priorities of packet
  - —Still under development
  - —See RFC 2460
- Flow Label
  - —Used by hosts requesting special handling
- Payload length
  - —Includes all extension headers plus user data

IP v6 Header Fields (2)

- Next Header
  - —Identifies type of header
    - Extension or next layer up
- Source Address
- Destination address

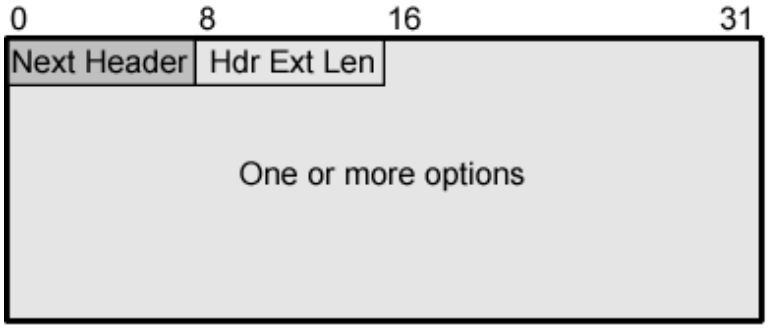
IPv6 Addresses

- 128 bits long
- Assigned to interface
- Single interface may have multiple unicast addresses
- Three types of address

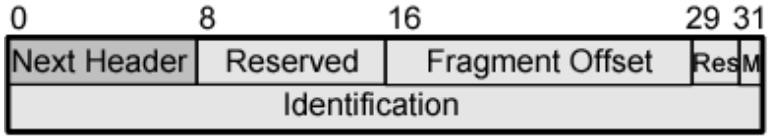
Types of address

- Unicast
  - —Single interface
- Anycast
  - —Set of interfaces (typically different nodes)
  - —Delivered to any one interface
  - —the “nearest”
- Multicast
  - —Set of interfaces
  - —Delivered to all interfaces identified

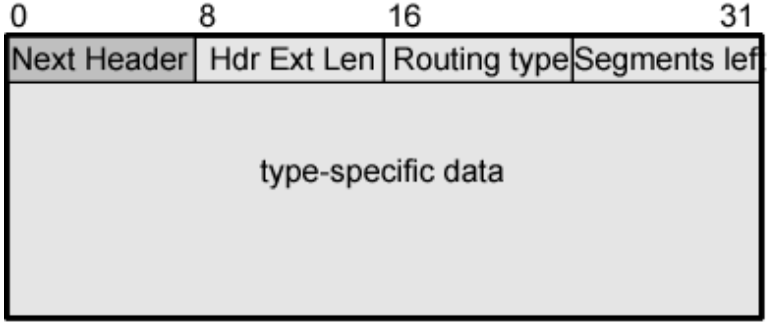
IPv6 Extension Headers



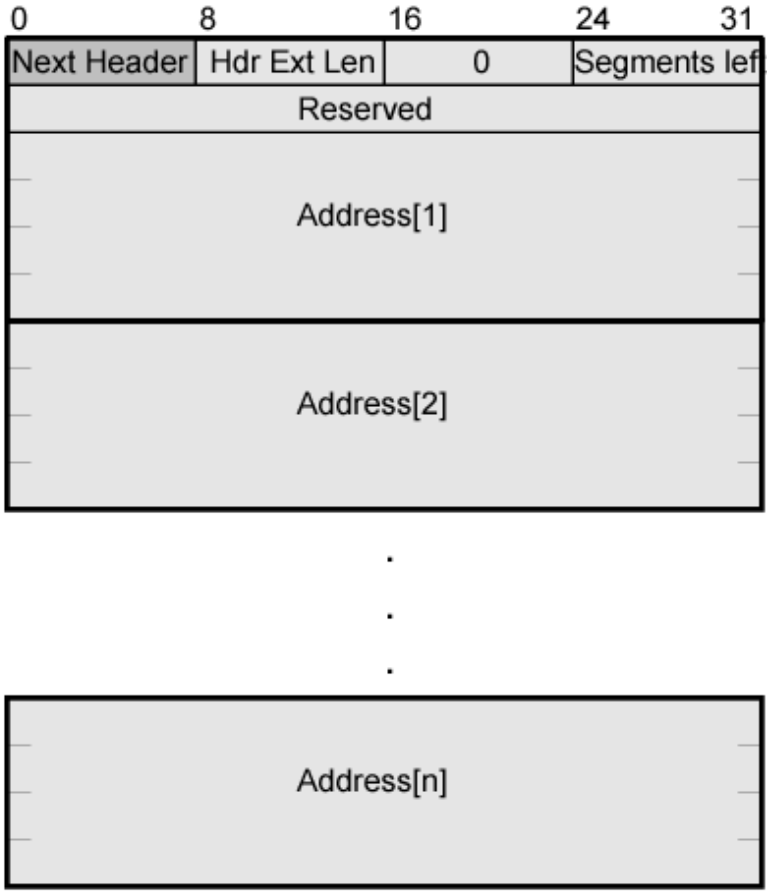
(a) Hop-by-hop options header; destination options header



(b) Fragment header



(c) Generic routing header



(d) Type 0 routing header

Hop-by-Hop Options

- Next header
- Header extension length
- Options
  - Pad1
    - Insert one byte of padding into Options area of header
  - PadN
    - Insert *N* (≥2) bytes of padding into Options area of header
    - Ensure header is multiple of 8 bytes
  - Jumbo payload
    - Over 2^{16} = 65,535 octets
  - Router alert
    - Tells router that contents of packet is of interest to router
    - Provides support for RSPV (chapter 16)

Fragmentation Header

- Fragmentation only allowed at source
- No fragmentation at intermediate routers
- Node must perform path discovery to find smallest MTU of intermediate networks
- Source fragments to match MTU
- Otherwise limit to 1280 octets

Fragmentation Header Fields

- Next Header
- Reserved
- Fragmentation offset
- Reserved
- More flag
- Identification

Routing Header

- List of one or more intermediate nodes to be visited
- Next Header
- Header extension length
- Routing type
- Segments left
  - i.e. number of nodes still to be visited

Destination Options

- Same format as Hop-by-Hop options header

Required Reading

- Stallings chapter 18
- Comer, S. *Internetworking with TCP/IP, volume 1*, Prentice-Hall
- All RFCs mentioned plus any others connected with these topics
  - www.rfc-editor.org
- Loads of Web sites on TCP/IP and IP version 6